Data Science Capstone Project

By Marc Hauptmann

A comparison of Chicago neighbourhoods in terms of business and social activity against crime rate.

Introduction / Business Problem:

Crimes rates in Chicago are higher than the US average, especially when it comes to violent crimes such as murder and rape [Ref1]. Various studies have been conducted and articles written about the potential causes of the peak in crime rate in the city in recent times (e.g. [Ref2]). Potential causes such a decline in educational or social spending or the reduction of police presence and activity on the streets, have been widely dismissed. Underlying social-economic causes are more challenging to assess, due to the complex interactions and / or only incomplete mapping of the underlying contributing factors.

This Data Science Project is an attempt at solving 2 questions of importance to the 'bigger' problem described above:

- Is there a relationship between density and types of business venue activity in Chicago neighborhoods and the frequency and type of crimes committed?
- How do neighborhoods in Chicago compare in terms of 'prosperity' and 'safety' (i.o.w. what are safe, flourishing neighborhoods in Chicago and what distinguishes them from 'languishing', 'crime ridden' ones)?

This is a question of importance for multiple stakeholders:

- Individuals and agencies active in the field of crime prevention and/or responsible for decisions on deployment of police force.
- Individuals and agencies (incl. city government) active in the area of neighborhood social and business development, who are responsible to plan and budget such projects based on need and potential outcome.
- Individuals and business, who need to decide where to locate and settle themselves in the Chicago area.

Data requirements and -sources:

For the proposed study, two types of data will be required: 1. A dataset that represents business venue activity per neighborhood in Chicago 2. A dataset that contains an overview of crimes committed in the Chicago area incl. the locations of the crimes committed.

Dataset #1 can be constructed out of various data sources: a) An overview of Chicago Neighborhoods incl. geospatial definitions of the neighborhood boundaries for visualization and data exploration purposes. Datasets are available in multiple formats such as GeoJSON from [Ref3]. b) Data on venues incl. types of venues, longitude and latitude, with the flexibility to collect said data for each neighborhood within a perimeter from the center of the neighborhood. Such data is available via Foursquare [Ref4].

Dataset #2 is based on an annual overview of all crimes committed in the Chicago area, incl. type of crimes, longitude and latitude and so forth. Such a dataset is provided via the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system [Ref5].

Both datasets will be processed, linked, and correlated to solve the questions posed in the previous section. The utilized analytical approach will be described in more detail in the next section.

[Ref1]: http://www.city-data.com/crime/crime-Chicago-Illinois.html, accessed 13-dec 2020.

[Ref2]: "What's Causing Chicago's Homicide Spike?" by Matt Ford, The Atlantic, January 25, 2017, https://www.theatlantic.com/politics/archive/2017/01/chicago-homicide-spike-2016/514331/ , accessed 13-dec 2020.

[Ref3]: Neighborhood boundaries in Chicago, developed by the Office of Tourism: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9 , accessed 13-dec 2020.

[Ref4]: Foursquare website: https://foursquare.com/ , accessed 13-dec 2020.

[Ref5]: https://www.chicago.gov/city/en/dataset/crime.html , accessed 13-dec 2020.